Homework assignments should be single spaced, 12 point font. You do not need to provide the course information in this document. Please include section headings (such as "Project Background") and feel free to also include subheadings where appropriate. The text should not be written in first person (for example, avoid saying "I will do this .. then I .."). Please be sure to check your grammar and spelling. Also, please feel free to use bullet points and bold text when appropriate (but use sparingly throughout document).

Data may be obtained from Kaggle (as discussed in the course materials), other publicly available datasets, or you may use your own data. Personal data may include data from work if appropriate. You may not repeat a data project/use data from another class. Multiple data sources may be used/merged for your project.

Homework 1:

Project Background, Data Background, Problem Formulation, and Data Strategy Plan (1 page)

Project Background: Provide a summary of the project and if appropriate the company and its mission. Provide a framework of your project and who it is meant for. Is it a company, government entity, someone you work for? In the simplest terms: what is the background that is going to set up the problem(s) you are addressing in this project.

The information for this section may include information from the company's website or the website where you obtained the data; however, please be sure to summarize the information in your own words and properly reference your sources. Direct quotations of information from a website should be minimal or nonexistent. This should be big-picture information of who is the client that your project will help and big-picture information about the data. Data specifics such as variables, timeframe will come in a later section.

Problem Formulation: This includes what question(s)/problem(s) are being addressed through your analysis. These should be clearly stated. The goal of the project should be discussed in a manner that is clear to a client (not a data scientist). For our purposes, assume that your client does not understand data science methods or terms, but does care about the question(s) being addressed and insights gained.

Data Strategy Plan: This goes hand-in-hand with the objective(s) of your project. HW4 details the requirements for the types of analyses you will be using in this project. At this point, you should begin to create a data strategy plan and how to utilize the data to provide the client with insights that help solve the problem(s) you are addressing.

Provide some background to your data. Is it all company-specific data? Does it include financial data, environmental data, etc.? Did the data come from multiple sources? If so, provide a brief background on those and how it connects to your project.

Topics in this section must include at minimum: What is the dataset you are using, how was it obtained, the timeframe the data covers, files utilized, what level is the data you are using (for example, are observations at the transaction, day, week, month, year, city, state, country etc). Lastly, identify your variable(s) of interest and how they relate to the objective of your project. These should be the variables that you think you will include your formal analysis and may include your target variable if appropriate to your analysis. (If you have a large number of variables you are using in your analysis, you can list the variables or refer to an entire file, but you must still connect them to the goal of your project.)

It is important to note that the plan that you lay out in this homework and problem you identify may change over the next week or two. That is okay. Digging into the data may result in modifications to this section. Your project document is fluid, so while you may have a plan submitted for this week's assignment, know that you are always able to modify your plan and in doing so update this section of your report. However, be careful to not get distracted by too many ideas or completely change your project at the last minute.

Homework 2:

Summary of Data Cleansing and Exploratory Data Analysis (1 additional page)

Summary of Data Cleansing and EDA: This includes a discussion of the important steps taken while processing the data and exploring the general structure of the data. For example, what assumptions did you make? Were there any missing (null or error) observations, and if so, how were they handled? What about outliers? Did you have to combine files? If so, did you have to make any decisions related to those steps (for example, an inner vs outer merge)? Did you use the full dataset provided (for example if there were multiple years, did you include them all)? Did you have to normalize variables or were there strings that you wanted to convert into dummy variables? What other steps did you take while cleansing the data? Remember that in this document we do not care about your code (do not refer to specific lines in your code) instead we care about business insights and how your decisions or assumptions relate to the insights gained.

Also, be sure to provide the client with some of the insights gained from this stage of your analysis. For example, you must include information such as the number of observations in your sample (maybe differences between starting sample size and final if some observations were removed), the mean/median, min, max of variables of interest, and visualizations of your data.

Again, the information in this document should be discussed in a manner that is clear to a client (not a data scientist). For our purposes, assume that your client does not understand data science methods or terms, but does care about the question(s) being addressed, general steps and methods that you took, and insights gained.

Homework 3:

Data Visualizations, Ethics and Future Tools (1-2 additional pages)

Provide at least two meaningful visualizations of your data. "A picture is worth a thousand words." This is especially true when providing the final report of your project to a client. The analyses that you do as part of homework 3 may also be considered exploratory, however, the visualizations included here should be separate for any visualizations associated with your analyses in homework 2. In addition to the visualizations, there should be titles and labels.

Include a section in your written document that details what insights can be drawn from your data visualizations. Lastly, visualizations should be labeled, for example: Figure 1, Figure 2 … and your written report should have text that references the figures and information. For example, you may say "Figure 1 shows ... and from this we find that …"

This unit you are also tasked with thinking about the ethical implications of your data and analysis. Please include one paragraph addressing the ethical considerations that you have made and any potential ethical implications that would be relevant to your analysis.

Lastly, while not a requirement, please feel free to use any new tools while completing your project. If you do you any tools that were not covered in your prior DATA courses, please include a short description of them as an Appendix to your written report.

Homework 4:

Analysis, Business Insights, and Final Summary (3-5 additional pages)

Analysis and Business Insights: Your project must include *at least two methods of analysis*. You may choose models from: regression analysis, cluster analysis, classification analysis, or recommender systems. You may choose to incorporate two different regression models, or you may choose two completely different methods of analysis.

For the analysis, state what methods you used, what variables are the focus in your analysis, what question does the analysis addresses, why you chose the method you used, what you expected to gain from this method of analysis, and provide a detailed discussion of the results. The results can also be depicted with charts, images, a confusion matrix, etc., but a written discussion should be included and accompany the images. This section should be the main portion of your written report. Be sure not only to present the results, but also explain what they mean and what insights can be gained.

An important component of this part of the report is the business insights that the client can gain from your analysis. The discussion should be presented in a way that the client can easily understand. As such, you *should not* show or discuss your code. Relatedly, when talking about the models, you do not need to go into technical details, but instead want to convey the information in a way that the client (not a data scientist) can easily understand. For our purposes,

assume that your client does not understand data science methods or terms, but does care about the insights gained.

Final Summary: You should conclude your report with a separate summary of your project. This includes restating your objective and concisely stating your findings.

For week 4 homework, you should start each of these sections; however, if they are not fully complete, that is okay. The final complete document is due with homework 5. For full credit for homework 4, you need at least 1 page about each model with the knowledge that you will have the full required length and conclusion section also included for the final document.

Homework 5:

Make revisions to Final Report (one file that includes all information for Homework 1-4) and submit Final Document on D2L

Make any necessary revisions to your final report and submit it to D2L before the deadline.

Submit peer presentation feedback

Provide peer feedback provided on other students' projects and presentations. I will provide a form for to complete for credit for this part of the assignment. Please upload your completed form to D2L.

Specifically, the form will ask the following: What was the objective of the student's project? What is one thing the student did well? What is one thing that could be improved?

Submit final Jupyter Notebook, and related files

Please upload your final Jupyter Notebook and files to D2L. If needed, please feel free to zip the files prior to uploading them to Dropbox. Please note, the data for your project must also be provided.